

*Advancing Science with DNA Sequence*

# Laboratory Science Program (LSP)



**Objective: To provide an early LSP overview  
and a perspective on applications to  
DOE missions**

# Laboratory Science Program Activities

## Site Visits

ORNL June 9  
LANL June 22  
PNNL July 24  
LBNL July 25

“Plant Genomes for Energy Applications” Advisory Board Meeting @ JGI June 21, 2006

POC Meeting @ ORNL July 5-6, 2006

LSP Overview and Critique  
Advisory Board Review and Recommendations  
Large Genome White Paper Reviews

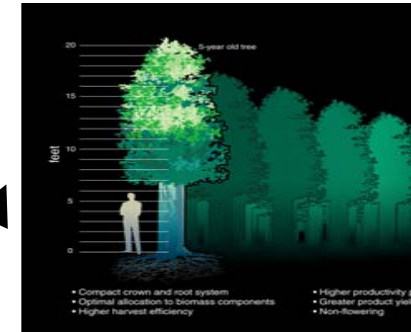
# Goals and Components of the LSP

## Goals

- Facilitate sequencing science at the DOE National Laboratories
- Develop cross-laboratory large-scale sequencing projects that advance missions
- Develop avenues for the service and outputs of PGF to meet the needs of the National Laboratories

## Components

- Small-scale Sequencing
  - Single PI
  - Supplemental Contributions
- Large-scale Sequencing
  - Multi-institutions
  - Targeted Species Selection



# Partners and Collaborators

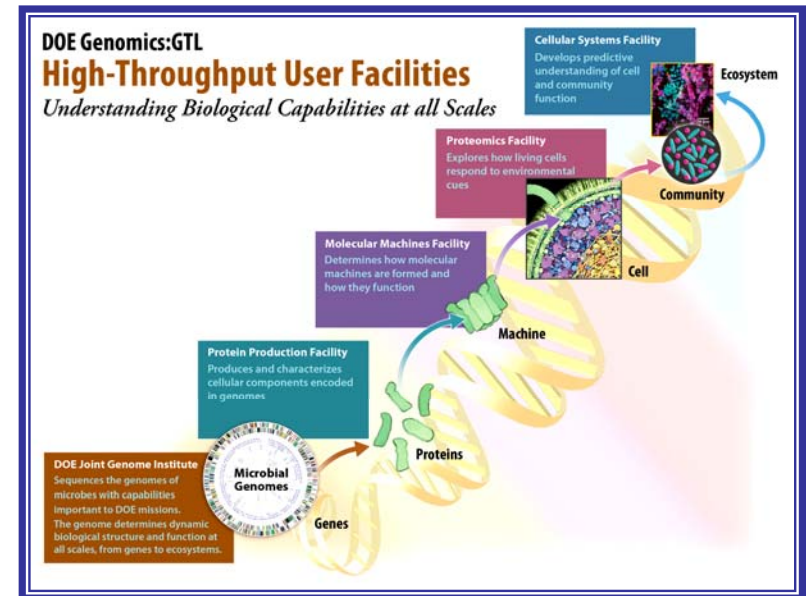
## Partner

## POC

- LANL Jim Brainard
- LBNL Joe Gray
- LLNL Elbert Branscomb
- ORNL Reinhold Mann
- PNNL Doug Ray

## Collaborators

- NREL
- ANL
- BNL



U.S. Department of Energy  
**Energy Efficiency and Renewable Energy**

# Small-scale Sequencing (S<sup>3</sup>) Component

- **Sequencing projects include:**
  - shotgun sequencing of individual BACs
  - microbial community metagenomes
  - PCR-amplified genomic segments
  - small-insert clone libraries & EST clone libraries
- **A total of approximately 1 Gb annually will be allocated among approximately 20 to 40 projects.**

# Small-scale Sequencing Projects

Principal Investigator	National Laboratory	Title	Requested (Mb)
C. Anderson	BNL	Paired-End Sequencing Tags to Identify p53 Binding and Methylated Loci in the Human Genome after Low Dose Radiation	7.0
G. Anderson	LBNL	Microbial Ecology of New Orleans Floodwaters	10.4
V. Bailey	PNNL	Towards a census of the microbial populations in Tallgrass Prairie soil	20.2
S. Baker	PNNL	Understanding <i>Aspergillus terreus</i> itaconic acid production	11.3
C.f Han	LANL	Sequencing Transcriptomes of <i>Acidothermus cellulolyticus</i> , <i>Frankia</i> sp. Strain EAN1pec, <i>Anaeromyxobacter dehalogenans</i> 2CP-C and <i>Dehalococcoide</i> sp. Strain BAV1	
C. Kuske-1	LANL	Genome sequencing of the five major C and N <sub>2</sub> -fixing bacteria in biological soil crusts from arid land soils	16.0
J. Magnussen	PNNL	Phylogeny of the Columbia River Microbiota	4.3
J. Pett-Ridge	LLNL		9.7
L. Stubbs	LLNL	Sequencing ChIP-enriched binding sequences for uncharacterized primate transcription factors	10.0



# Small-scale Sequencing Projects

## How to you submit a proposal?

**Use the Community Sequencing Program web application form.**

<http://www.jgi.doe.gov/programs/LSP/s3callforprops.html>

## When should you submit a proposal?

**Whenever you have an idea ready to go.**

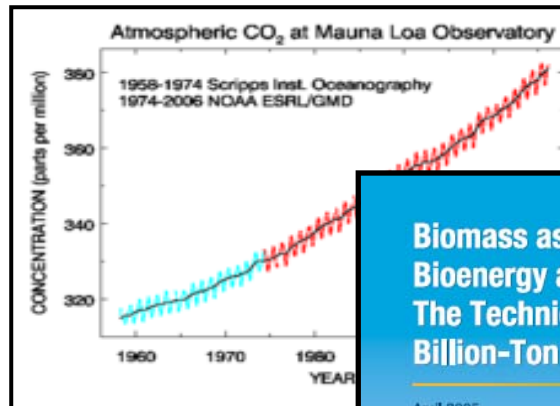
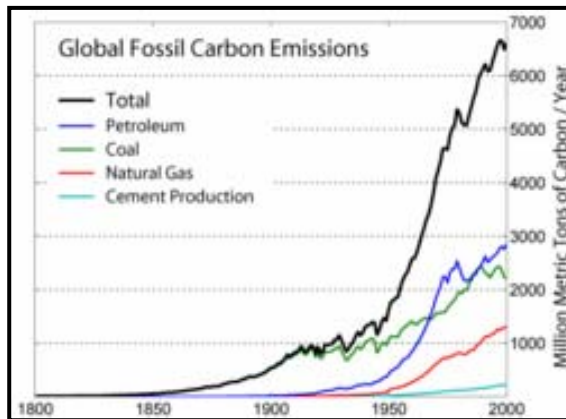
# Low Dose Subcomponent

- **An 18 member, multi-lab panel was convened and 3 research-area recommendations were made:**
  - radiation-induced cancers to identify genome regions that confer individual susceptibility
  - cells response to high vs. low doses
  - sequencing genomic regions of known susceptibility genes



# Bioremediation

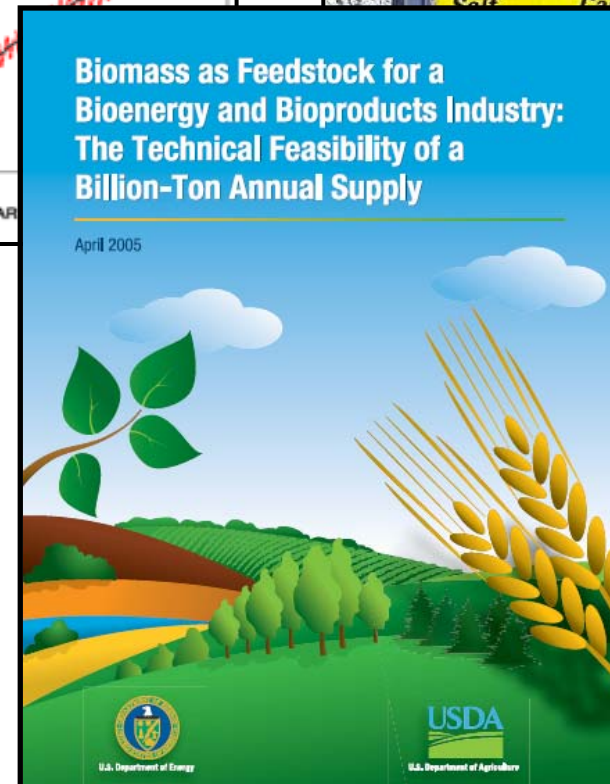
# Large-scale Sequencing Component -- Rationale



DOE: Energy Information Administration  
<http://www.eia.doe.gov/>

## Solution – Biomass Feedstocks

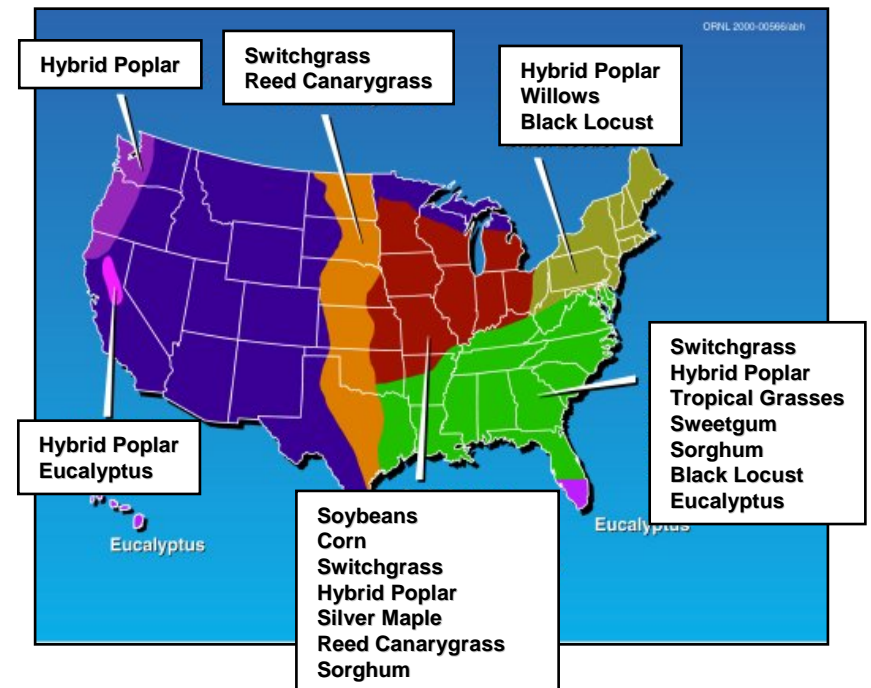
- 10 dry tons / acre / year
- \$0.73 / gallon of ethanol
- 20-30 million acres



[http://www1.eere.energy.gov/biomass/pdfs/final\\_billionton\\_vision\\_report2.pdf](http://www1.eere.energy.gov/biomass/pdfs/final_billionton_vision_report2.pdf)

# Large-scale Sequencing Component -- Rationale

- **Near-term deployment of plant-based systems to address issues related to biofuels development & atmospheric carbon mitigation** will require the application of **genome-based science and tools**.



- Sequencing project will focus on 10-15 moderately-sized plant genomes that have direct links to DOE missions and/or have a phylogenetic position that make their genomes informative.

# Large Genome Sequencing Efforts

## **“Fungal Genomes for Energy Applications” Advisory Board**

**Scott Baker, Lead, PNNL**

**Tom Brettin, Fred Brockman, Mike Himmel, Diego  
Martinez, Mike Miller, Chris Schadt, Tamas Torok**

## **White Papers**

### **“Fungal Genome Sequencing for Bioenergy”**

#### **First Four Targets:**

<b>Cochilobolus heterstrophphus</b>	<b>35 Mb</b>
<b>Tremella mesenterica</b>	<b>20 Mb</b>
<b>Allomyces macrogynus</b>	<b>30 Mb</b>
<b>Theilevia terrestris</b>	<b>40 Mb</b>

# Large Genome Sequencing Efforts

## “Plant Genomes for Energy Applications” Advisory Board

**Joe Ecker**

**Salk Institute**

**Jeff Dean**

**University of Georgia**

**Dan Cullen**

**University of Wisconsin**

**Bert Abbott**

**Clemson University**

**Steve DiFazio**

**West Virginia University**

**Michael Freeling**

**UC Berkeley**

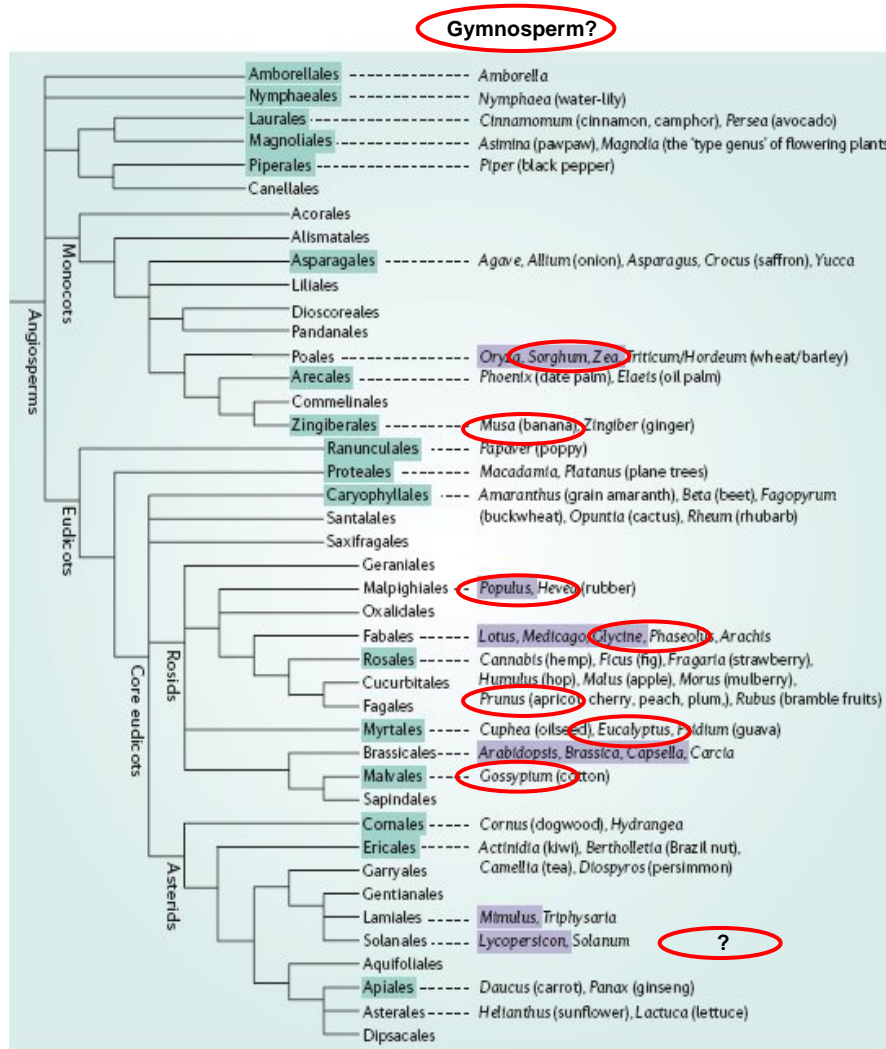
## White Papers

**“Plant Genomes for Bioenergy”**

**Targets TBD**



# Large-scale Sequencing Approach Plant Genomes



Andrew Paterson (2006) Nature Reviews Genetics

## Proposed Criteria

- DOE Relevance
- Informative position
- Small - moderate sized genome
- An active user community
  - Nat. Lab.
  - Academic
  - International
- Genetic & genomics resources

 -- proposed

 -- current activity



# Populus as an Large-scale Sequencing Example

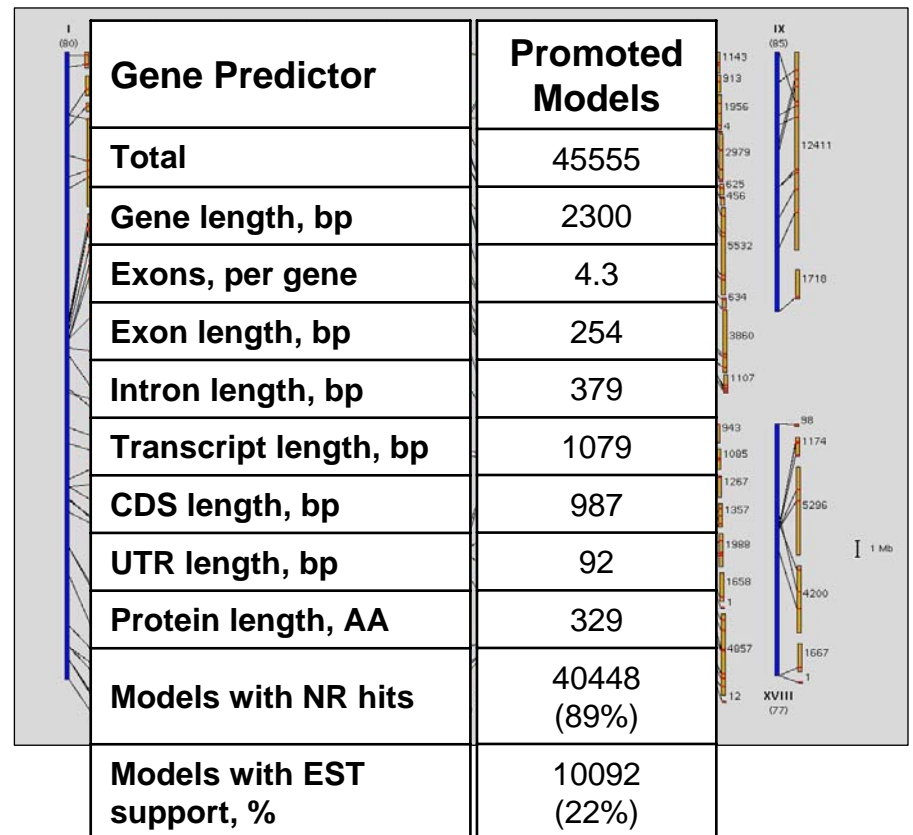
- *Populus trichocarpa*
  - 'Nisqually-1'
  - Female
  - Located in 1995
  - Nisqually river in Washington
- Shot-gun cloning and sequencing using 3 alternate insert libraries
  - 2.5-3.0 kb
  - 6.5-8.0 kb
  - 36 kb
- 7.2 million sequence reads
- ~4.2 billion high-quality bases
- 7.5X coverage of 480 $\pm$ 10 Mb genome



# Populus as an Large-scale Sequencing Example

## Assembly Statistics

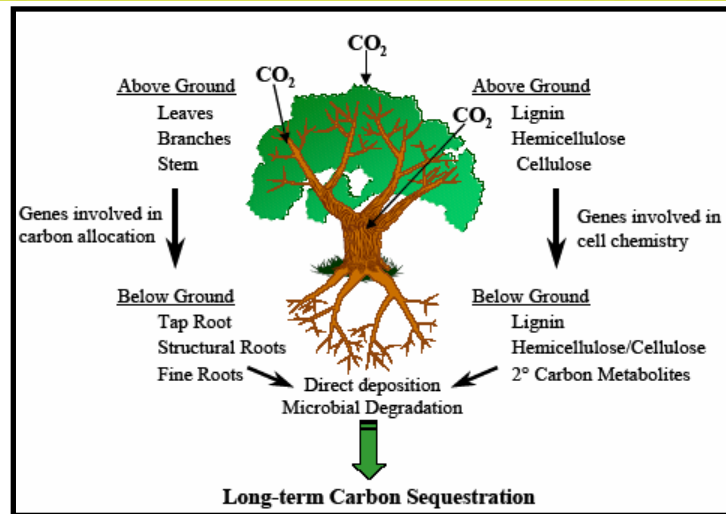
- Total number of scaffolds: 2447
- Total length of scaffolds: 429 Mb
- N50 scaffold number: 58
- N50 scaffolds size: 1.9 Mb
- Total number of contigs: 15154
- Total length of contigs: 334 Mb
- N50 contig number: 690
- N50 contig size: 127 kb



Gene Predictor	Promoted Models
Total	45555
Gene length, bp	2300
Exons, per gene	4.3
Exon length, bp	254
Intron length, bp	379
Transcript length, bp	1079
CDS length, bp	987
UTR length, bp	92
Protein length, AA	329
Models with NR hits	40448 (89%)
Models with EST support, %	10092 (22%)

# Populus as an Large-scale Sequencing Example

Bioenergy



Global  
Climate  
Change

Carbon  
Sequestration

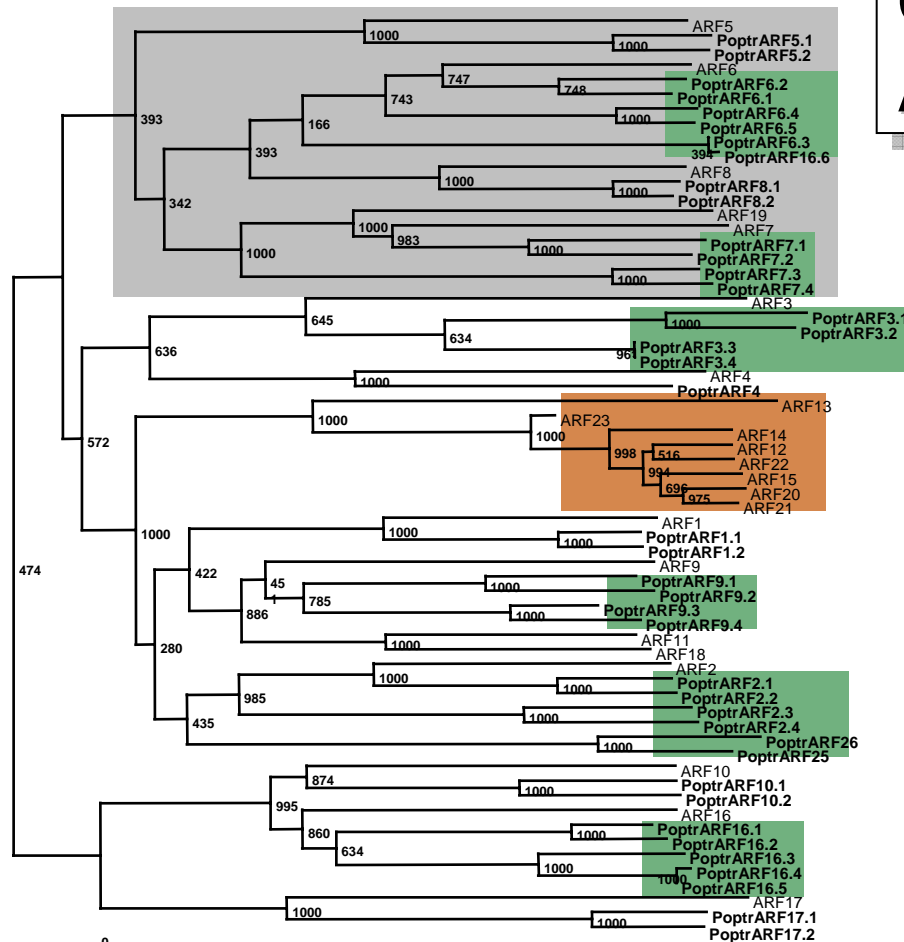
Phytoremediation

Understand **allocation and partitioning** of photosynthates (carbon).

Understand roles of Aux/IAA and ARFs in determining **architecture** of plant (root).

Identify genes/QTLs controlling **root phenolics** (residence time).

Identify **accelerated domestication** genes (drought resistance).



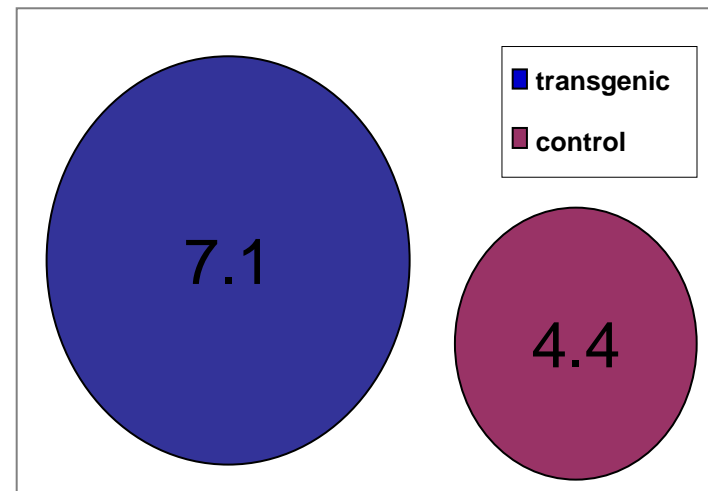


# Populus as an Large-scale Sequencing Example

90-day-old *Populus* cuttings

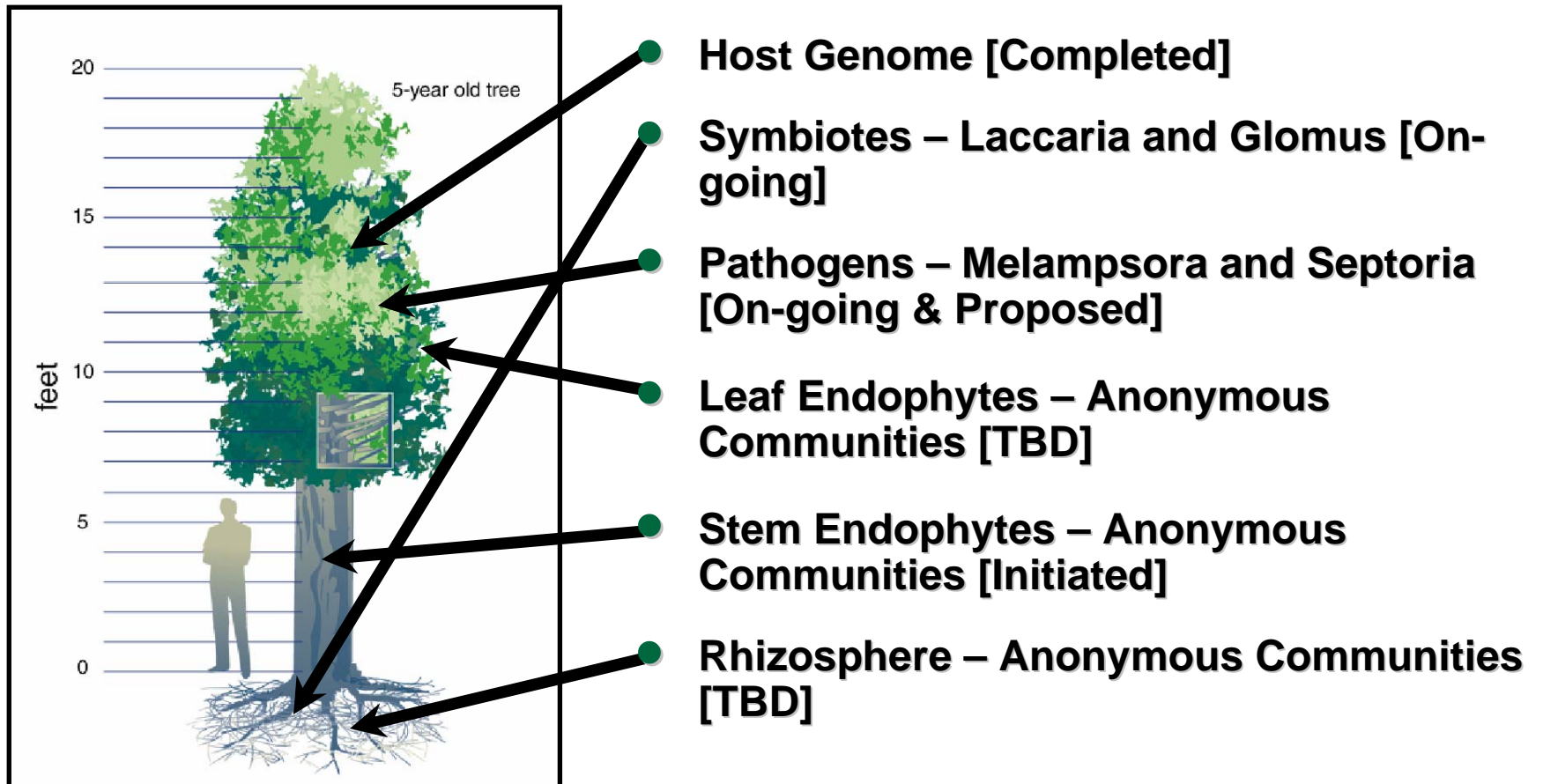


Enhanced radial growth of IAA16.3 transgenics vs. controls



stem cross sectional area (cm)

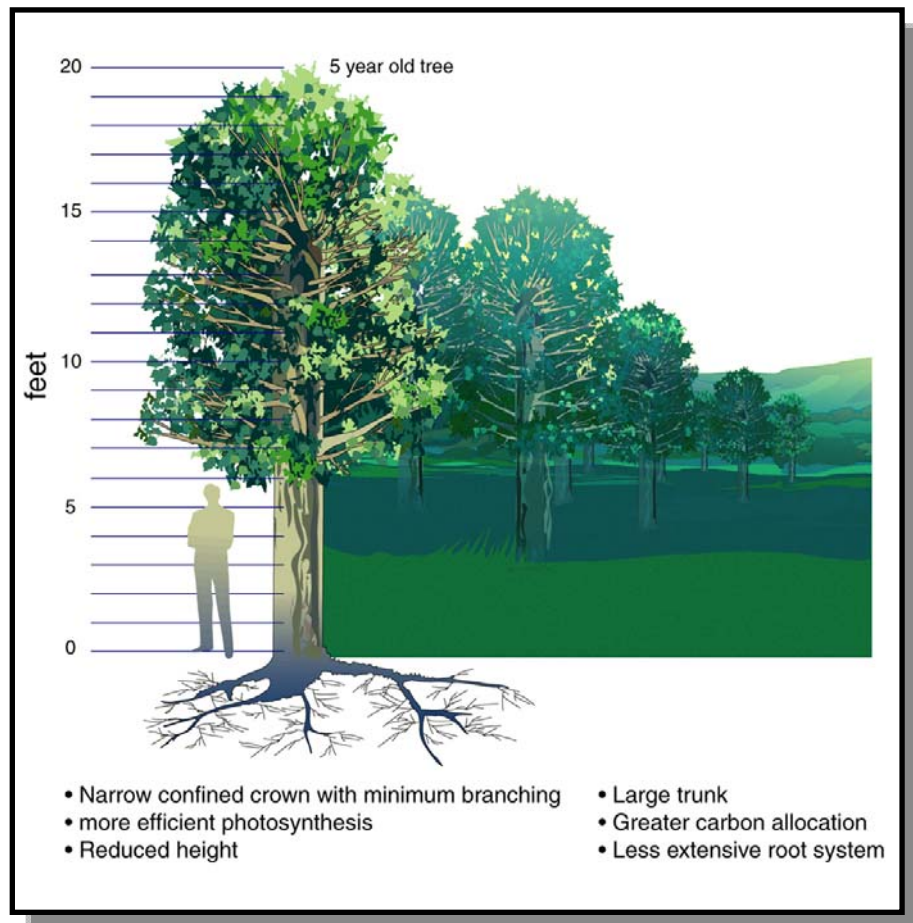
# Biomass/Biofuels Metagenome



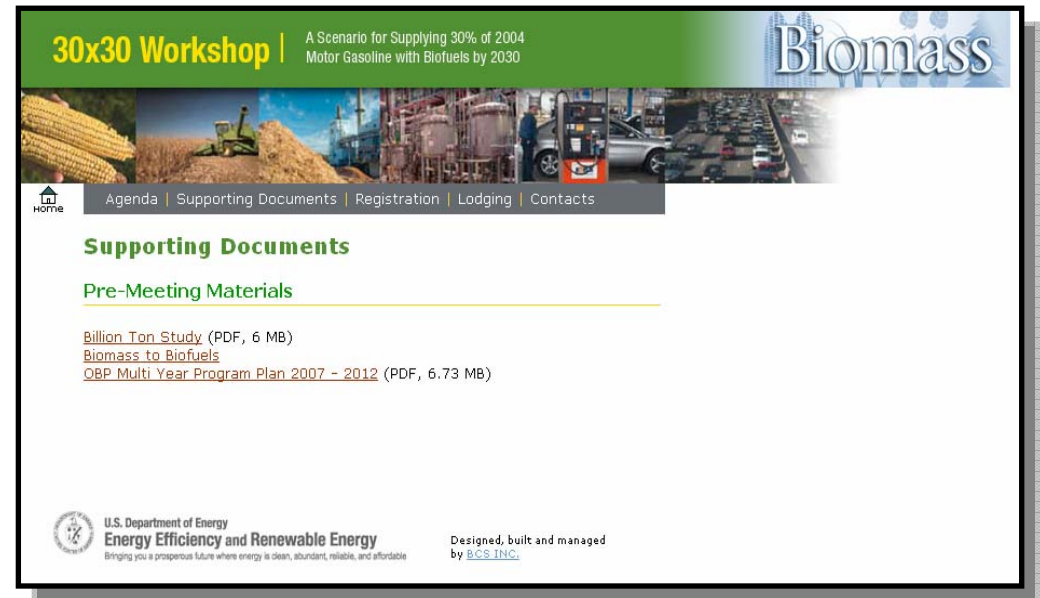
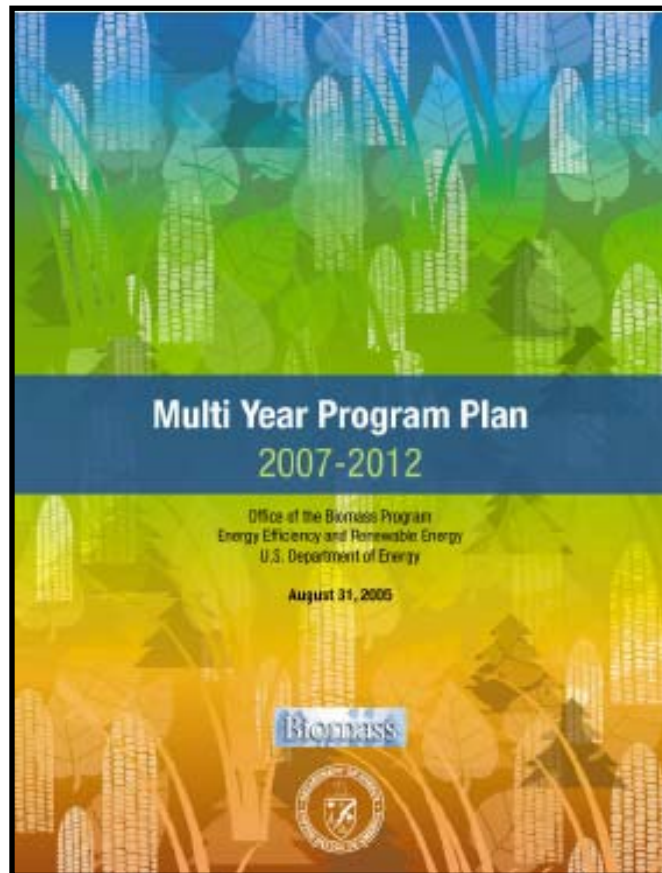


# Summary

- **The Laboratory Science Program will facilitate inter-laboratory science projects involving small scale and large multi-institution sequencing efforts**
- **A genomics-based systems biology approaches to DOE relevant missions will accelerate the development of new sources of energy and chemical feedstocks**



# Can we grow these programs?



## FY07 requests –

**\$4.5M Regional Feedstock Partnerships**

**\$5.5 M Feedstock Infrastructure Development**

<http://www1.eere.energy.gov/biomass/pdfs/mypp.pdf>